Correlation between IoT sensor measurements and total electricity consumption in smart homes

Marko Jelić[*a], Dea Pujić[*a], Marko Batić [a]

[*] School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11120 Belgrade, Serbia

[a] Institute Mihajlo Pupin, University of Belgrade, Volgina 15, 11060 Belgrade, Serbia

{marko.jelic, dea.pujic, marko.batic}@pupin.rs

***Abstract*—With the IoT industry rapidly growing and evolving, smart home concepts powered by distributed IoT sensors are becoming more and more popular amongst residential users. This transformation has created an opportunity for novel energy management solutions with the main goal of improving energy efficiency. A key element in this process is being able to predict energy consumption with this paper aiming to extend the current state of the art by analyzing real measurements obtained from smart sensors through their distribution, statistical characteristics and suitability as predictors in a forecasting algorithm.**

I. Introduction

The concept of IoT-powered households (so-called “smart homes”) is slowly entering the mainstream with more and more users expressing interest in living in a connected home. This rapidly growing concept, currently valued as an industry at 212 billion USD worldwide and with a projected value of 1 trillion USD [1], intertwined with the ongoing transition in the telecommunication sector from for 4G to 5G [2] networks which promises the lower latencies and higher bandwidths, could potentially be one of the most interesting fields for researchers, engineers and residents in the coming future. With the current number of connected IoT devices currently placed at around 27 billion, this number is expected to grow to over 75 billion by 2025 [3], [4], [5]. Having these numbers in mind, the main question that should be posed is what will be the best ways to utilize the expansion of IoT in a world that will grow more connected by the day. This undeniable potential has been recognized by the European Commission through the foundation of the Alliance for Internet of Things (AIOTI) [6] but also various organizations worldwide like the Internet of Things Consortium (IoTC), [7], IoT World Alliance [8], the Industrial Internet Consortium [9] and many others that are actively working on promoting and implementing various IoT solutions for both business and residents.

In the IoT landscape, another important player, besides various organizations and alliances, are the companies working on development and production of IoT devices. In smart homes, these devices can be used for different purposes such as energy management, improved comfort and lighting, home entertainment, security, health issue monitoring, control of smart appliance, etc. These devices are often augmented by either accompanying smartphone [10] or web apps or assistant voice assistants (like Google Assistant, Siri or Amazon Alexa) that can monitor and actuate upon them when the appropriate command is given with the examples of architecture backing these systems proposed in [11] and [12].

The introduction of IoT sensors in people’s homes has provided a wide variety of opportunities for new applications using the collected data. In the context of a research project, several houses in the vicinity of Leers, France, have been equipped by a set of sensors from a Danish manufacturer called *Develco Products* [13]. These sensors include:

- external metering interfaces (EMIs) for measuring total instantaneous power draw (demand) and integral energy consumption of the entire household;
- smart plugs and cables for measuring instantaneous power draw (demand) and integral energy consumption of individual appliances;
- smart thermostats which measure heat energy flow (demand) and integral energy consumptions of heating elements that use fluids;
- windows/door (open/close) sensors that also measure temperature;
- motion sensors which measure movement using passive infrared (PIR) sensors and estimate the occupancy (expressed by either 0 or 1) using a set of timeouts and PIR reading, but also measure temperature and illuminance;
- air quality sensors which measure volatile organic compound (VOC) levels, humidity, but also measure temperature.

Using these devices, user behavior can be analyzed in great detail, far more that was previously possible due to the fact that these IoT sensors provide key insight into internal measurements within the household in real time. One possible application of user behavior monitoring focuses on energy efficiency by simultaneously tracking current consumption in comparison to what can be estimated using a demand forecaster using historical data. In order to create such a forecaster, the suitability of different predictors has to be assessed and, in this regard, this paper aims to extend the research presented by [14] where external predictors were predominantly used (tariff indication, hour of day, day of week, holiday indication and previous consumptions). Through the introduction of IoT measurements from within the household (occupancy and indoor temperature as well as previous energy consumptions), this paper aims to analyze their suitability as predictors for the purpose of demand forecasting.

II. METHODOLOGY

From the pilot site, a set of 10 houses with the most stable connection between the gateway and the installed sensors were selected for the study conducted in this paper. The data from these houses is queried in three discrete time periods in three different seasons (between August 15th and 18th 2019, between October 15th and 18th 2019 and between December 15th and 18th 2019) and for each hour the following set of data is extracted (note that consumption only refers to electricity):

- consumption during that hour (*consumption*, [kW]);
- consumption of the previous hour relative to that hour (*consumption_previous_hour*, [kW]);
- consumption of the previous 24h relative to that hour (*consumption_previous_day,* [kW]);
- consumption of the previous 7 days relative to that hour (*consumption_previous_week,* [kW]);
- average total occupancy during that hour (*avg_occupancy*. [ ])
- average indoor temperature during that hour (*indoor_temperature_avg*, [°C])

The collected data is processed so that only measurement instances where none of these values are invalid remain. Furthermore, for each of the measured variables X, its mean value μ_X and standard deviation σ_X are estimated as

$$\bar{\mu}_X = {\sum_{i=1}^{N} X(i)} \Big/ {N}, \quad \bar{\sigma}_X^2 = {\sum_{i=1}^{N} (X(i) - \bar{\mu}_X)^2} \Big/ {N-1}.$$

When these values are determined, outlier removal is performed by excluding all measurement instances where one of the measurements from that hour falls outside of the $[\bar{\mu}_X - 2\bar{\sigma}_X, \bar{\mu}_X + 2\bar{\sigma}_X]$ range. After this filtering is performed, the distribution of the remaining data is presented in form of a normalized histogram with 50 bins, with the mean and standard deviations of the remaining data denoted in the upper right corner. The distribution of *consumption is* shown in Figure 1 while distirbutions for *consumption_previous_hour*, *avg_occuapncy and indoor_temperature_avg* illustrated in Figure 2. These datasets are analyzed as a whole but are also separated into seasonal sets with solely summer, solely fall and solely winter data in order to observe any seasonal variations that exist within the data.

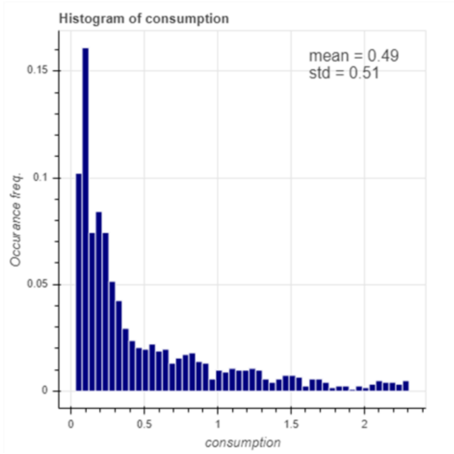

Figure 1. Distribution of *consumption* masurement

After the data distribution is analyzed, the measurements are analyzed crosswise in order to observe if they exhibit any correlation between each other. The resulting scatter plots, accompanied with a "best fit" straight line obtained using the least square algorithm from a Python library called *scipy.stats.linregress* [15] with the slope and intercept values denoted in the legend, are presented in Figure 3. In this figure, an example of the relationship between *consumption* plotted on all three y axes, and *consumption_previous_hour, avg_occupancy* and *indoor_temperature_avg* plotted on the respective x axes is given.

Having in mind that the residents in the analyzed apartments utilize electric space heaters and have no cooling equipment, in order to illustrate how seasonality impacts the results, scatterplots showing the relationship between *consumption* and *consumption_previous_hour*, *consumption* and *avg_occupancy* as well as *consumption* and *indoor_temperature_avg* are illustrated in Figure 4, 5 and 6, respectively.

Following the graphical analysis of the aforementioned gathered measurements, their correlation is calculated using the Pearson correlation coefficient ρ_{XY}. For two random variables X and Y, this coefficient can be estimated using the expression

$$\bar{\rho}_{XY} = {\sum_{i=1}^{N} \left((X(i) - \bar{\mu}_X)(Y(i) - \bar{\mu}_Y) \right)} \Big/ {\bar{\sigma}_X \bar{\sigma}_y}.$$

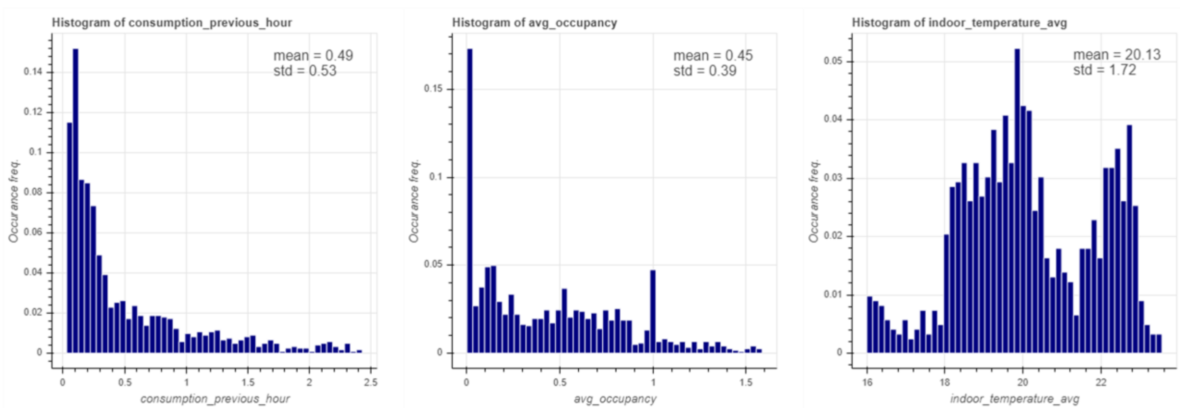

Figure 2. Histograms illustrating distributions of different IoT measurements

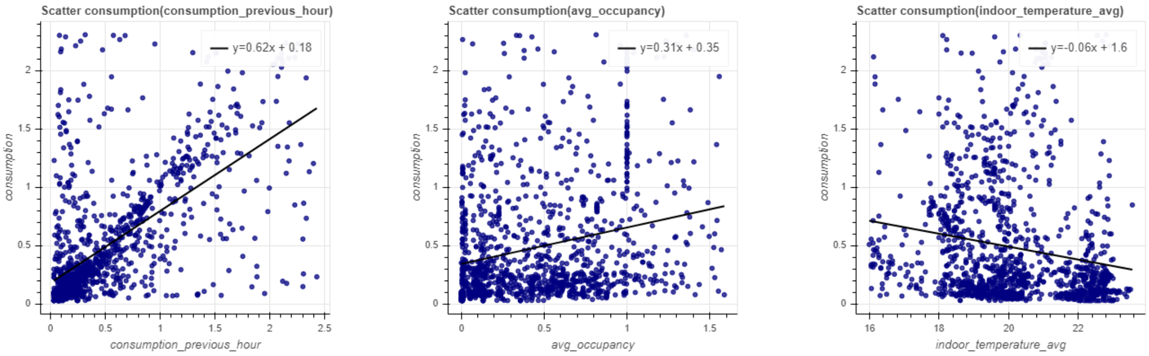

Figure 3. 2D scatterplots illustrating the relationship between consumption and other IoT measurements for full datasets (all three seasons combined, from left to right: *consumption_previous_hour, avg_occupancy and indoor_temperature_avg*)

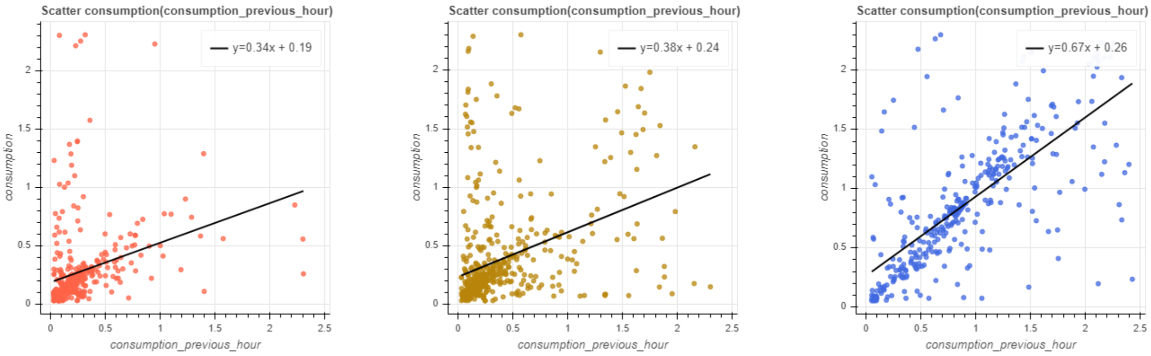

Figure 4. 2D scatterplots illustrating seasonal variation in relatonship between *consumption* and *consumption_previous_hour* (from left to right: summer, fall and winter)

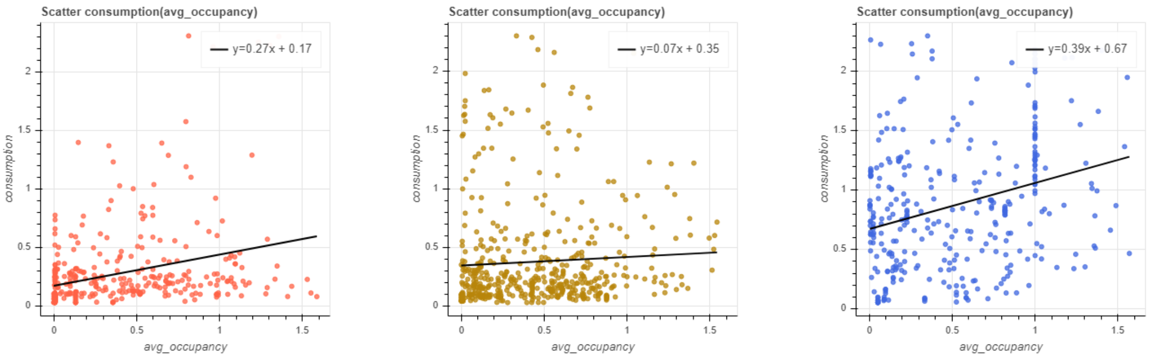

Figure 5. 2D scatterplots illustrating seasonal variation in realtionshib between *consumption* and *avg_occupancy* (from left to right: summer, fall and winter)

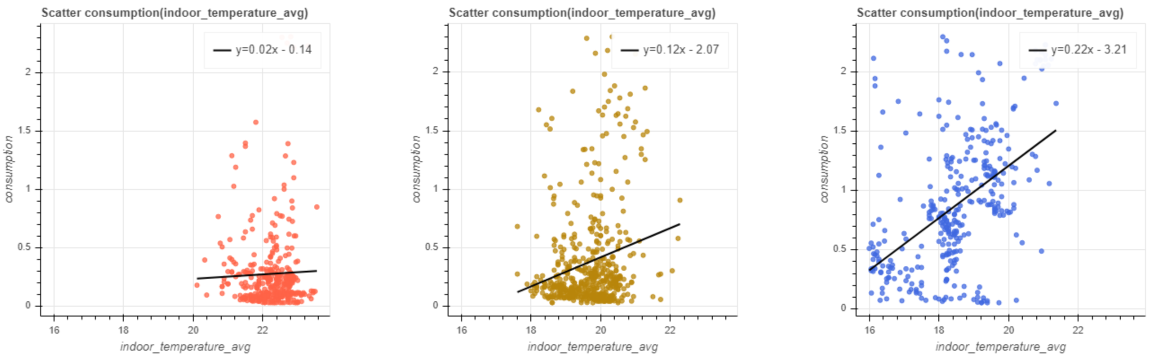

Figure 6. 2D scatterplots illustrating seasonal variation in relationship between *consumption* and *indoor_temperature_avg* (from left to right: summer, fall and winter)

TABLE I.
HEATMAP OF CORRELATION COEFFICIENTS BETWEEN DIFFERENT IOT MEASUREMENTS AND CONSUMPTION

	all	summer	fall	winter
consumption	1.000	1.000	1.000	1.000
consumption_previous_hour	0.630	0.306	0.375	0.710
consumption_previous_day	0.680	0.504	0.659	0.510
consumption_previous_week	0.558	0.257	0.064	0.490
avg_occupancy	0.236	0.305	0.058	0.289
indoor_temperature_avg	-0.186	0.032	0.214	0.498

III. RESULTS AND DISCUSSION

Out of the set of all measurements, one by one variable is selected and inserted into the formula for the correlation coefficient alongside *consumption*, as the consumption measurement is the one that is supposed to be estimated in demand forecasting applications. By doing this, the results from Table I are obtained with the given values only representing the cross-correlation coefficient value (i.e. the value of $\bar{\rho}_{XY}[1,2] = \bar{\rho}_{XY}[2,1]$) between the denoted variable in the leftmost column and *consumption*. As shown for the whole dataset, the highest correlation is exhibited between the current consumption and, interestingly, the consumption of the entire previous day (with a value of 0.68), closely followed by the previous hour consumption (0.63) and previous week consumption (0.56). Furthermore, occupancy also appears to be a suitable predictor in some cases (with a value of 0.24), albeit noticeably less correlated with current *consumption* than previous consumptions. This value could potentially be even higher if a different set of houses is analyzed due to the fact that the households from the specific pilot site in this study employ a set of ripple control devices coupled with time-of-use (ToU) tariff which results in a significant portion of the load being shifted to the night time when occupancy readings are expected to be lower since no motion is detected when the inhabitants are asleep and hence occupancy readings (which are derived from motion) equal zero during most of the night.

As indoor temperature is concerned, regarding the whole dataset in aggregate seems to blur the specific characteristics of seasonal temperature distributions resulting in a negative correlation with the consumption. However, when data for individual seasons is analyzed, all correlation coefficients turn positive, with an especially high correlation observable in winter time, showing that indoor temperature can also be used as a suitable predictor. This is especially true during cold portions of the year when high powered heating devices tend to be used more frequently. A similar result could have been the case in summer time if the considered households had any cooling devices. However, this was not the case and hence there is almost no correlation between *indoor_temperaturea_avg* and *consumption* during summer. Finally, for reference, if an average outdoor temperature variable is added with the help of an external data provider [16] for the considered location, the correlation coefficient with consumptions is estimated at -0.45 for the whole dataset and -0.20, -0.10 - 0.10 for summer, fall and winter respectively.

IV. CONCLUSION AND FUTURE WORK

This paper analyzes the relationship between different measurements reported by IoT sensors in several smart homes in France, specifically with regard to their influence on total hourly electric energy consumption. The results obtained provide an essential basis for the feature section procedure, necessary in development of demand prediction applications, through a correlation test between different IoT measurements that can be considered for this purpose such as previous consumptions, occupancy and indoor temperature measurements. Furthermore, the results show that, due to the presence of heating devices, seasonality, especially with regards to indoor temperature, plays a key role that effects the influence of different variables on the consumption.

ACKNOWLEDGMENT

The research presented in this paper is partly financed by the European Union (H2020 LAMBDA project, GA. No: 809965 and H2020 InBetween project, GA No. 768776), and partly by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] Shanhong Liu, “Internet of Things (IoT) - Statistics & Facts,” *Statista*. https://www.statista.com/topics/2637/internet-of-things/ (accessed May 29, 2020).

[2] S. Borkar and H. Pande, “Application of 5G next generation network to Internet of Things,” in *2016 International Conference on Internet of Things and Applications (IOTA)*, Jan. 2016, pp. 443–447, doi: 10.1109/IOTA.2016.7562769.

[3] “Smart Home Statistics [2020]: Growth of Connected Devices,” *iPropertyManagement.com*. https://ipropertymanagement.com/research/iot-statistics (accessed Jan. 15, 2020).

[4] “Mapping Internet of Things innovation clusters in Europe,” Jun. 19, 2019. https://ec.europa.eu/digital-single-market/en/internet-of-things/clusters (accessed Jan. 15, 2020).

[5] “37 Impressive IoT Statistics: 2019 & 2020 Data Analysis & Market Share,” *Financesonline.com*, Sep. 04, 2019. https://financesonline.com/iot-statistics/ (accessed Jan. 15, 2020).

[6] "HOME," *AIOTI SPACE*. https://aioti.eu/ (accessed Jan. 15, 2020).

[7] "IoTC," *Internet of Things Consortium*. https://iofthings.org/ (accessed Jan. 15, 2020).

[8] "IoT World Alliance." http://iotworldalliance.org/ (accessed Jan. 15, 2020).

[9] "Industrial Internet Consortium." https://www.iiconsortium.org/ (accessed Jan. 15, 2020).

[10] L. Berbakov, M. Batić, and N. Tomašević, "Smart Energy Manager for Energy Efficient Buildings," in *IEEE EUROCON 2019 -18th International Conference on Smart Technologies*, Jul. 2019, pp. 1–4, doi: 10.1109/EUROCON.2019.8861616.

[11] L. Berbakov, N. Tomasevic, and M. Batić, "Architecture and Implementation of IoT System for Energy Efficient Living," in *2018 26th Telecommunications Forum (TELFOR)*, Nov. 2018, pp. 1–4, doi: 10.1109/TELFOR.2018.8611888.

[12] M. Domb, "Smart Home Systems Based on Internet of Things," *Internet of Things (IoT) for Automated and Smart Applications*, Feb. 2019, doi: 10.5772/intechopen.84894.

[13] "Develco Products - White label products for smart home, smart energy, home security, and assisted living." https://www.develcoproducts.com/ (accessed Jan. 15, 2020).

[14] A. Dedinec and A. Dedinec, "Correlation of variables with electricity consumption data," in *ICIST 2016 Proceedings*, 2016, pp. 118–123, Accessed: Jan. 15, 2020. [Online]. Available: http://www.eventiotic.com/eventiotic/library/paper/29.

[15] "scipy.stats.linregress — SciPy v1.4.1 Reference Guide." https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.linregress.html (accessed Jan. 15, 2020).

[16] "Weatherbit | Weather API Documentation." https://www.weatherbit.io/api (accessed Jan. 15, 2020).